UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/837,158	04/19/2001	Karen Mae Holland	ARC000018US1	8446

48146 7590 07/11/2005

MCGINN & GIBB, PLLC
8321 OLD COURTHOUSE ROAD
SUITE 200
VIENNA, VA 22182-3817

EXAMINER	
BLACKWELL, JAMES H	

ART UNIT	PAPER NUMBER
2176	

DATE MAILED: 07/11/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/837,158	HOLLAND ET AL.	
	Examiner	Art Unit	
	James H. Blackwell	2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *26 April 2005*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-12 and 14-23* is/are pending in the application.

4a) Of the above claim(s) *13* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-12 and 14-23* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *19 April 2001* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. This Office Action is in response to Amendment received 04/26/2005.

2. Claims 1-12, and 14-23 are pending. Claims 1, 17, and 23 are independent claims.

3. Claim 13 has been cancelled.

4. The objection to claims 12, 14, and 21 has been addressed in part, by converting Claim 14 to independent form. However, in lieu of a general review of the Clustering art, and an update of the prior art search, such objections have been withdrawn. These claims have now been rejected.

5. Claims 1-12, and 14-16 were rejected under 35 U.S.C. 101 as being directed to non-statutory subject matter.

Claim Rejections - 35 USC § 112

6. The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

7. Claims 1, 12, and 14-21 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the enablement requirement. The claim(s) contains subject matter, which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. Specifically, the use of the term "structured variable(s)". In the Specification, the Applicant suggests as examples that structured variables are time

intervals (pg. 1), public opinion (pg. 3), predetermined time intervals (days, weeks, months) (pg. 4, 10), but never specifically defines the term in such a way as to distinguish it from multiple interpretations by the skilled artisan.

8. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

9. Claims 1, 12, and 14-21 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Specifically, the meaning of the term "structured variable(s)". It is unclear from the Specification what the meaning of structured is, nor is it clear what the term variable refers to, rendering both the individual terms and the combination of the two indefinite.

Claim Rejections - 35 USC § 103

10. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-6, 8, 11, 17-18, and 21-23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Allan et al. (hereinafter Allan, "Topic Detection and Tracking Pilot Study Final Report", Proc. of DARPA Broadcast News Transcription and Understanding

Workshop, 02/1998) in view of Goldszmidt et al. (hereinafter Goldszmidt, "*A Probabilistic Approach to Full-Text Document Clustering*", 1998, Technical Report ITAD-433-MS-98-044, SRI International).

In regard to independent Claim 1 (and similarly independent Claims 17, and 23), Allan teaches *a computer-implemented method for identifying relationships between text documents and structured variables pertaining to said text documents* by discussing a technique used by a group at Carnegie Mellon Univ. to detect events (news stories) from a corpus of documents (Secs. 3, 3.2).

Allan describes discovery of natural patterns of news stories over concepts (lexicon terms) and time (Sec. 3.2, 1st paragraph).

Allan also describes a conventional vector space model for incremental clustering (*forming categories of said text documents using said dictionary and an automated algorithm*). A story is presented as a vector whose dimensions are the stemmed unique terms in the corpus (*generating a dictionary of keywords in said text documents*).

Allan also teaches *counting occurrences of said structured variables, said categories and combinations of said structured variable and said categories combinations for said text documents* in that the calculation of term weighting in a story vector combining the within-story term frequency (TM) and the Inverse Document Frequency (IDF) (Sec. 3.2, 2nd paragraph).

Allan fails to explicitly teach *calculating probabilities of occurrences of said combinations of structured variables and categories to identify a relationship between said text documents and said structured variables*. However, Goldszmidt teaches a similarity measure based on probability (an overlap measure), which measures the degree of overlap between pairs of documents (p. 4, Sec. 2, Equation 1). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Allan and Goldszmidt as both documents discuss aspects of clustering techniques. Adding Goldszmidt provides the benefit of computing probabilities to measure similarity.

In regard to dependent Claim 2, Allan teaches that *said algorithm comprises a keyword occurrence algorithm and wherein each of said categories comprises a category of text documents in which a particular keyword occurs* in that the use of a clustering algorithm using a vector space model. Clusters are constructed based on the similarity of vector spaces containing similar concepts (lexicon terms) and similar times (Sec. 3.2, 1st paragraph).

In regard to dependent Claim 3, Allan teaches that *said algorithm comprises a clustering algorithm and wherein each of said categories comprises a category of said text documents containing a particular cluster* in that the use of a clustering algorithm using a vector space model. Clusters are constructed based on the similarity of vector spaces containing similar concepts (lexicon terms) and similar times (Sec. 3.2, 1st paragraph).

In regard to dependent Claim 4, Allan teaches *said clustering algorithm comprises a k means algorithm* using a cosine similarity measure (Sec. 3.2, 3rd paragraph) with clustering. K-means using a cosine similarity measure is often called spherical k-means. Hence, one can infer that Allan uses a k-means clustering algorithm.

In regard to dependent Claim 5, Allan teaches *said forming said categories comprises inputting a predetermined number of categories* in the use of a k-means clustering algorithm (see analysis in Claim 4), and it is notoriously well known that at the heart of a k-means clustering method is the input of a predetermined number of clusters (categories are cluster names).

In regard to dependent Claim 6, Allan teaches *said forming said categories comprises: generating a sparse matrix array containing a count of each of said keywords in each of said text documents* in that the use of a clustering algorithm using a vector space model. Clusters are constructed based on the similarity of vector spaces containing similar concepts (lexicon terms) and similar times (Sec. 3.2, 1st paragraph). It is also notoriously well known in the art to implement a set of vectors such as those taught by Allan into a sparse matrix for the purpose of evaluation of the corpus of documents.

In regard to dependent Claim 8, Allan fails to teach *said calculating probabilities comprises using a Chi squared function*. However, Goldszmidt teaches using a Chi-Squared test as part of the analysis of clustering methods (p. 15, 3rd paragraph). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Allan and Goldszmidt as both documents discuss aspects of clustering techniques. Adding Goldszmidt provides the benefit of using statistical measures to analyze clustering results.

In regard to dependent Claim 11, Allan teaches that *said generating a sparse matrix array comprises: third parsing text in said text documents to count a number of times that each of said keywords occurs in each of said text documents* in that the use of a clustering algorithm using a vector space model. Clusters are constructed based on the similarity of vector spaces containing similar concepts (lexicon terms) and similar times (Sec. 3.2, 1st paragraph). It is also notoriously well known in the art to implement a set of vectors such as those taught by Allan into a sparse matrix for the purpose of evaluation of the corpus of objects. Allan does not elaborate on a method for creating the sparse matrix, however it is notoriously well known that such a matrix would contain the number of times that each of a lexicon of terms occurred in the corpus of objects to be clustered.

In regard to dependent Claim 12, Allan fails to teach that *said relationships comprise said combinations of structured variables and categories having a lowest probability of occurrence*. However, it is notoriously well known in the art that measures of whether or not two objects are grouped together or not depend on how closely or how distant characteristics of two objects are in comparison to one another. Those that are distant in terms of their similarities would translate to having a low probability of occurrence. Likewise, such similarity measures would also allow one to deduce how likely the clustering of two objects is the result of randomness.

In regard to independent Claim 14, Claim 14 reflects the method for identifying relationships between text documents and structured variables pertaining to said text documents as claimed in Claim 1 (and similarly Claims 17, and 23) and Claim 12, and is rejected along the same rationale.

In regard to dependent Claim 18, Allan does not specifically teach *a memory for storing occurrences of said structured variables, categories and structured variable/category combinations and probabilities of occurrences of said structured variable/category combinations*. However, it would have been obvious to one of ordinary skill in the art at the time of invention to assume that such data would have to have been stored on some media such as memory, disk, or other computer storage, providing the benefit of ready access to the data for processing on a computer.

In regard to dependent Claim 22, Allan teaches that *said relationships comprise statistically significant relationships* the use of a clustering algorithm using a vector space model. Clusters are constructed based on the similarity of vector spaces containing similar concepts (lexicon terms) and similar times (Sec. 3.2, 1st paragraph). At the heart of clustering is determining relationships between objects. Similar objects are grouped together based on a similarity measure of some sort. Allan teaches the use of a standard cosine similarity test (p. 34, 3rd Column, 2nd paragraph). It is notoriously well known that similarity measures are typically a combination of statistical measures.

11. Claims 7, 9-10, 15-16, and 19-20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Alan in view of Goldszmidt and in further view of Yang et al. (hereinafter Yang, "Learning Approaches for Detecting and Tracking News Events", 1999, IEEE Intelligent Systems).

In regard to dependent Claim 7, Allan fails to specifically teach that *said keywords comprise at least one of words and or phrases, which occur a predetermined number of times in, said text documents*. However, Yang teaches that each document is represented by a vector of weighted terms that can be either words or phrases (p. 34, 2nd column, 4th paragraph). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Allan, Goldszmidt and Yang as all three deal with clustering issues related to object comparison and grouping. Yang's teaching provides the benefit of further elaborating on the summary taught by Allan.

In regard to dependent Claim 9, Allan fails to specifically teach that *said generating a dictionary of keywords comprises: first parsing text in said text document to identify and count occurrences of words; storing a predetermined number of frequently occurring words; second parsing text in said text documents to identify and count occurrences of phrases; and storing a predetermined number of frequently occurring phrases.* However, Yang teaches that each document is represented by a vector of weighted terms that can be either words or phrases (p. 34, 2nd column, 4th paragraph). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Allan, Goldszmidt, and Yang as all three deal with clustering issues related to object comparison and grouping. Yang's teaching provides the benefit of further elaborating on the summary taught by Allan.

In regard to dependent Claim 10, Allan teaches that *said frequently occurring words and phrases are stored in a hash table* in that he use of a clustering algorithm using a vector space model. Clusters are constructed based on the similarity of vector spaces containing similar concepts (lexicon terms) and similar times (Sec. 3.2, 1st paragraph). It is also notoriously well known in the art to implement a set of vectors such as those taught by Allan into a sparse matrix for the purpose of evaluation of the corpus of objects. It is also notoriously well known to store vectors, matrices in hash tables to enable their efficient storage and subsequent evaluation on a computer.

In regard to dependent Claim 15 (and similarly dependent Claim 19), and dependent Claim 16 (and similarly dependent Claim 20), Allan describes *said structured variables comprise predetermined time intervals* and *said predetermined time intervals comprise one of days, weeks, months and years* in the discovery of natural patterns of news stories over concepts (lexicon terms) and time (Sec. 3.2, 1st paragraph). Hence, Allan's teaching utilizes time as a structured variable for determining which cluster a given news story object belongs in. In addition, Yang suggests using time intervals in the evaluation of similarity of news story events (p. 34, 1st Column, bulleted paragraphs). In addition, Fig 1ab depicts the number of stories detected over time in days. It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Allan, Goldszmidt, and Yang as all three deal with clustering issues related to object comparison and grouping. Yang helps to further define time intervals.

In regard to dependent Claim 21, Claim 21 reflects the method for identifying relationships between text documents and structured variables pertaining to said text documents as claimed in Claim 14, and is rejected along the same rationale.

Conclusion

12. Any inquiry concerning this communication or earlier communications from the examiner should be directed to James H. Blackwell whose telephone number is 571-272-4089. The examiner can normally be reached on Mon-Fri.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather R. Herndon can be reached on 571-272-4136. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306 (**after July 15th, new Fax number is 571-273-8300**).

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

James H. Blackwell
07/05/05

William L. Bashore
WILLIAM BASHORE
PRIMARY EXAMINER
7/7/2005